AFOSR-TR-97

0081

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
		FINAL REPORT 15 Sep 92 - 14 Sep 96

4. TITLE AND SUBTITLE
DEVELOPMENT OF FREE-SPACE OPTOELECTRONIC NEURAL SYSTEMS BASED ON HYBRID SI/PLZT TECHNOLOGY

5. FUNDING NUMBERS
62712E
7013/51

6. AUTHOR(S)
Professor Sadik C. Esener

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, San Diego
La Jolla, CA 92093

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NE
110 Duncan Avenue Suite B115
Bolling AFB DC 20332-8050

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
F49620-92-J-0520

11. SUPPLEMENTARY NOTES

19970210 011

12a. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

On this grant research was conducted in three directions to make progress towards the goal of implementing large scale learning neural networks on compact optoelectronic multichip modules. (1) Refinement of the Fuzzy ARTMAP neural network algorithm. The algorithm is very well suited to hardward implementation, but previously did not perform well on most real problems. (2) Design, fabrication, and characterization of birefringent computer generated holograms to implement the OTIS optical system for reflective modulators. This optical system provides a k-shuffle interconnection pattern that may be used to impelement a backpropagation neural network. (3) Development of a single, compact, efficient optical system (OTIS-VE) compatible with surface-normal transmitters (such as VCSELs) implementing the k-shuffle interconnection pattern.

14. SUBJECT TERMS			15. NUMBER OF PAGES
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	

Final

~~Annual Status~~ **Report**

for

"Development of Free-Space Optoelectronic Neural Systems Based on Hybrid Si/PLZT Technology"

Sponsored by

ARPA Under Contract No. F49620-92-J-0520

Grantee

The Regents of the University of California

University of California, San Diego

La Jolla CA 92093

Reporting Period: 3/15/95 - 9/14/96

Principal Investigator: Sadik C. Esener
(619) 534-2723

Program Manager: Dr. A. Craig
(202) 767-4931

Development of free-space optoelectronic neural systems based on Hybrid Si/PLZT technology

ARPA contract #F49620-92-J-0520 Final ~~Annual~~ **report (11/1/96)**

Reporting period: ~~3/15/95 - 9/14/96~~ 9/15/92 - 9/14/96

Table of contents

1. Objectives and status

The objective of this program has been to develop optoelectronic technologies, components, and subsystems for the implementation of large-scale learning neural networks that can be constructed of compact optoelectronic multi-chip modules. Neural networks belong to a distinctive class of computing paradigms in that they naturally exhibit parallelism and alleviate the control task by learning from examples. Our goal is to demonstrate these key features in a hardware implementation.

During this reporting period, we conducted research in three directions to make progress towards that goal:

- Refinement of a the Fuzzy ARTMAP neural network algorithm. The algorithm is very well suited to hardware implementation, but previously did not perform well on most real problems.
- Design, fabrication, and characterization of birefringent computer generated holograms to implement the OTIS optical system for reflective modulators. This optical system provides a *k*-shuffle interconnection pattern that may be used to implement a backpropagation neural network.
- Development of a simple, compact, efficient optical system (OTIS-VE) compatible with surface-normal transmitters (such as VCSELs) implementing the *k*-shuffle interconnection pattern.

2. Fuzzy ARTMAP modifications for intersecting class distributions

The Fuzzy ARTMAP neural network performs a supervised clustering operation. It may be applied to vector quantization or classification of data, including speech, images, and radar range profiles. The algorithm is very well suited to implementation in parallel hardware[1]. However, as currently defined, the algorithm performs poorly due to cluster proliferation when the class distributions of the training data intersect, as commonly occurs with real-world data. We propose several modifications to the algorithm which eliminate the cluster proliferation problem. The performance of the original and modified algorithms is illustrated using the Peterson-Barney vowel data[2].

2.1. Cluster proliferation

Fuzzy ART[3] is a clustering algorithm that operates on vectors with analog-valued elements. Adding a further layer of processing (the "MAP field") to Fuzzy ART yields a *supervised* clustering algorithm, Fuzzy ARTMAP[4]. During learning, Fuzzy ARTMAP is presented with training vectors which have been labeled according to the class to which each belongs. For example, vectors of features extracted from short segments of speech might be labeled according to the phoneme being spoken during each segment. The Fuzzy ART module assigns the current training vector to a cluster. The MAP field associates a class with each cluster. If the class label of the training vector differs from the class associated with the chosen cluster, the MAP field causes the Fuzzy ART search to be repeated, this time with attention restricted to smaller clusters. If this procedure, termed *match tracking*, successfully assigns the training vector to a cluster with the correct class association, the weights defining that cluster are updated, making it more likely to capture that training vector in the future. If no cluster with the correct class association is found, a new cluster is created.

Classification of points from intersecting class distribution has long been identified as a more difficult problem than classification of points from separable distributions[5]. It is not possible to classify a point lying in the region of overlap with 100% certainty. The best a system can do is estimate the probability

that a training point belongs to a particular class, or equivalently determine a fuzzy value of the class membership of the data point[6]. The output of Fuzzy ARTMAP is restricted to a single class label. In this case, the best generic approach is to select the class from which the data point was most probably drawn. Intersecting distributions are common in real problems due to noise, classification uncertainty and errors, use of sub-optimal features, loss of context, etc. An example of such a distribution occurring with real-world data is shown in Figure 1.

When clouds of data points associated with several classes overlap, Fuzzy ARTMAP tends to create about one cluster per data point. We refer to this phenomenon as cluster proliferation. This is a direct result of the fact that the classes are not separable and that Fuzzy ARTMAP will create a new cluster whenever the training point can not be assigned to the correct class by match tracking. While this results in nearly perfect classification of the training vectors, performance on the test data (i.e. generalization) is often worse than if fewer clusters had formed. Furthermore, the time required for the Fuzzy ART search and the memory required to store the weights are both proportional to the number of clusters, providing additional incentives to limit the number of clusters created.

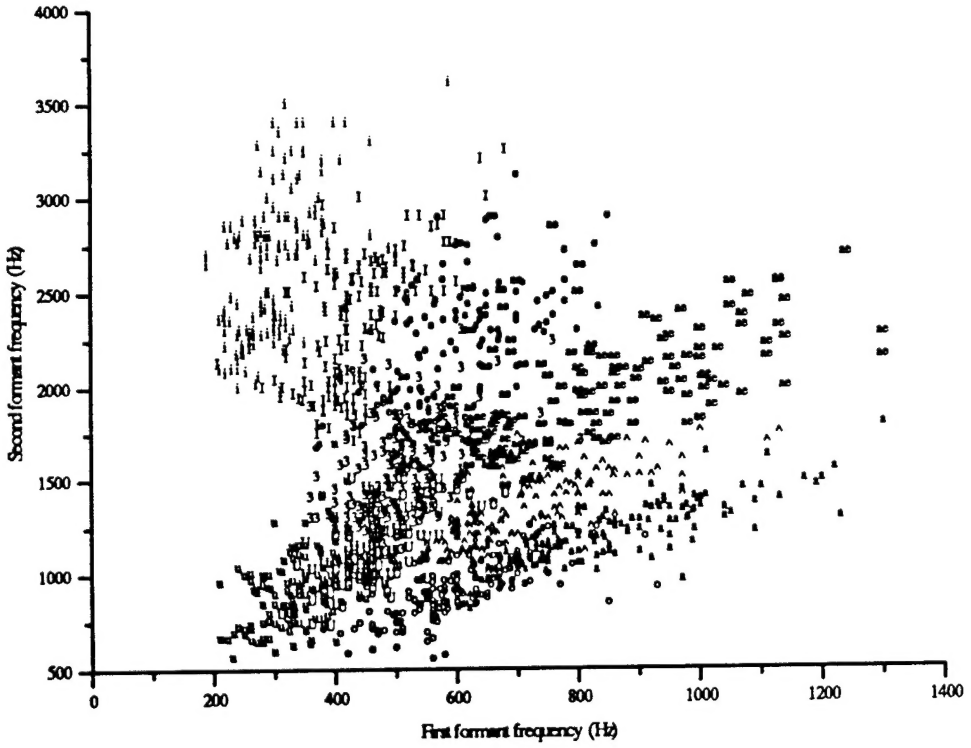

Figure 1 The ten American vowels, spoken by 76 speakers. Each utterance is plotted as a point in a two dimensional space defined by the first and second formant frequencies. The class distributions are strongly overlapping.

2.2. Standard Fuzzy ARTMAP algorithm

Fuzzy ART clusters vectors based on two separate distance criteria, *match* and *choice*. For input vector I and cluster j, the match function is defined by

$$S_j(I) \equiv \frac{|I \wedge w_j|}{|I|} ,$$

where w_j is an analog-valued weight vector associated with cluster j. $\wedge$ denotes the fuzzy AND operator, $(p \wedge q)_i \equiv min(p_i, q_i)$, and the norm $|\cdots|$ is defined by $|p| \equiv \sum_i |p_i|$. The choice function is defined by

$$T_j(I) = \frac{|I \wedge w_j|}{\alpha + |w_j|} \tag{1}$$

where α is a small constant. Increasing α biases the search more toward clusters with large $|w_j|$. Fuzzy ART assigns each input vector to the category that maximizes $T_j(I)$ while satisfying $S_j(I) \geq \rho$ (the *match criterion*), where the *vigilance*, ρ, is in the range $0 \leq \rho \leq 1$. *Unused clusters* ("clusters" associated with unused weight vectors) always satisfy the match criterion and have a constant choice value, T_{unused}.

The MAP field may be envisioned as a set of N_{clu} processing elements (PEs), each with N_{cat} weights, where N_{clu} is the number of clusters formed by the Fuzzy ART module and N_{cat} is the total number of classes (categories). It is basically used as an N_{clu} by N_{cat} look-up table. During training, if cluster J wins the Fuzzy ART competition and MAP field entry $(J, class_J)$ of the table is not greater than some threshold, ρ^{ab}, ρ is increased and the Fuzzy ART search is repeated. If MAP field entry $(J, class_J)$ is greater than ρ^{ab}, the data point is considered to have been correctly classified. Typically, one weight of each MAP field PE equals *1* and the others equal *0*, associating a single class with each cluster, and $\rho^{ab} = 0$.

Once the training vector has been assigned to the correct class, the weights of the winning cluster are updated by

$$w_{Ji}^{new} = \begin{cases} w_{Ji}^{old} & w_{Ji} \leq I_i \\ w_{Ji}^{old} - \beta\left(w_{Ji}^{old} - I_i\right) & w_{Ji} > I_i \end{cases},$$

where $0 < \beta \leq 1$.

Creating a new cluster: If the training vector is assigned to an unused cluster, the weight vector of that cluster is set equal to the training vector and the associated MAP field weights are adjusted to reflect the training vector's class.

2.3. Modifications to Fuzzy ARTMAP

2.3.1. Pruning

If those clusters capturing the fewest training vectors are periodically removed, the number of clusters in use approaches a steady state value. Clusters that are not removed can grow to capture training vectors previously in the pruned clusters. If clusters capturing less than N_{prune} training vectors are periodically removed, the absolute maximum number of remaining clusters is N_{prune} / N_{train}, where N_{train} is the total number of training data points. In practice, the number of clusters formed is usually significantly smaller than this.

A simple way to keep track of how many training points are captured by each cluster is to change slightly how the MAP field weights are updated. Instead of setting a weight to *1* when a new cluster is formed, the weight is incremented by some constant, δ_{hit}, every time a training vector is assigned to that cluster. The size of the weight now indicates how many training vectors the associated cluster has captured. In order to

prevent the MAP field weights from growing beyond bounds, a time decay term is added. We implement this by dividing all MAP field weights by 2 at the end of every iteration through the training data.

2.3.2. Restricted creation of new clusters

The rate of creation of new clusters may be limited. For example, if the same data set is presented repeatedly during training, the number of clusters that is created during each iteration may be limited to some small value, for example the total number of classes. This causes existing clusters to grow as much as possible, generally reducing the total number of clusters formed.

2.3.3. Choice function favoring larger clusters

As described by Carpenter, et al.[4], the Fuzzy ART weights associated with each cluster may be envisioned as two points defining a hyper-rectangle in M-dimensional space, where M is the number of components in each data vector (before complement coding). As a result of the original *Weber law* choice function (eq. 1), if a data point falls inside both of two overlapping rectangles, the smaller rectangle (corresponding to the larger $|w_j|$) wins. This guarantees that a cluster created in response to some training vector captures the training vector if it is immediately presented again (unless another rectangle with zero area already exists at the same position, as may happen when the training set contains two identical vectors with different class associations). Furthermore, if a data point falls midway between two weight rectangles, the smaller one wins. While it is necessary to preserve the former property, the latter one has undesirable consequences. In particular, a new cluster will capture data points from a large volume of feature space previously captured by a nearby large cluster. Intuitively, a small weight rectangle results from a very localized set of training points, whereas a large weight rectangle results from a spread out set of training points. If a test point lies equidistant from the two rectangles, it lies more standard deviations away from the smaller rectangle than from the larger one, and therefore should be assigned to the latter. The following alternate choice function provides the desired classification:

$$T_j^L(I) \equiv |w_j|\left(1 + \left[|I \wedge w_j| - |w_j|\right]/\alpha\right)$$

As in the Weber law choice function, α is a small constant. Here, α determines the range at which a larger weight rectangle is favored over a smaller one. $T_j^L(I)$ may be less than 0. Note that the quantity $|w_j|$ need only be computed (and stored locally) when the weights of cluster j are updated.

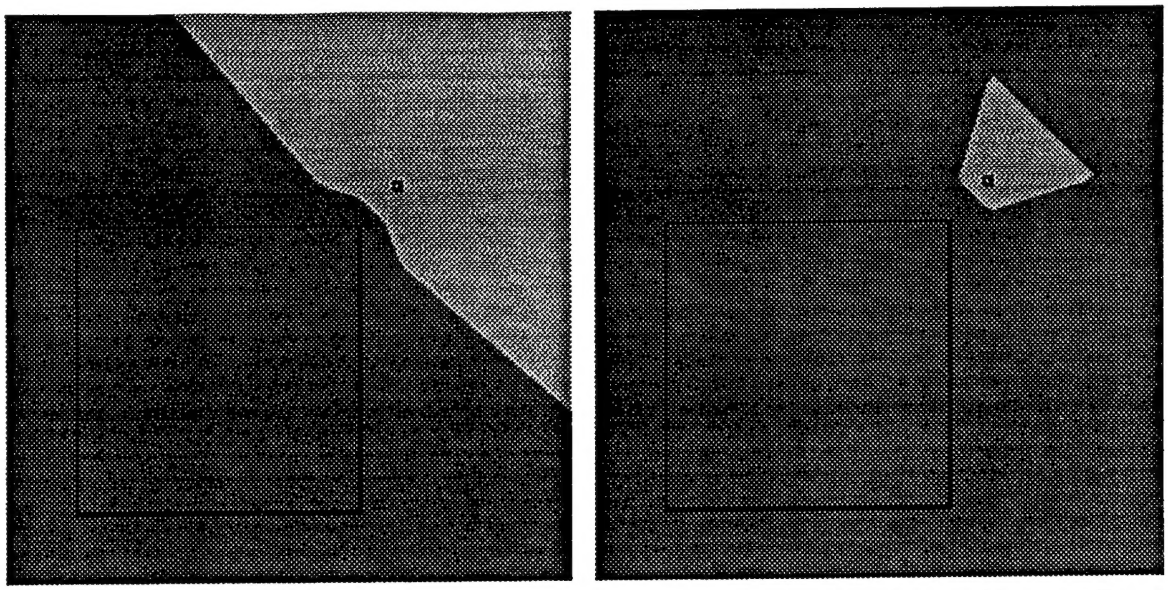

Figure 2 Win regions of two clusters, one with a large weight rectangle and one with a small weight rectangle, when the Weber law choice function is used (left) and when the alternate choice function favoring large weight rectangles is used (right).

2.3.4. Bi-directional weight movement

The Fuzzy ART learning law assures that weight rectangles only grow with time and never become smaller. Consequently, after enough iterations through the training data, every training point will lie either within or on the boundary of the weight rectangle of the cluster which captures it. This tends to create rectilinear win region boundaries throughout the feature space, which is often not desirable. In each dimension of the feature space, if the closer of the two points defining the weight rectangle is moved toward the training data point regardless of whether the point lies inside or outside the rectangle, the weight rectangle can grow and shrink. Thus the cluster boundaries can better approximate the shapes of the Bayesian decision boundaries between intersecting class distributions. Note that without pruning or limited rate of creation of new clusters, this method often leads to the creation of more clusters. One of the methods for limiting the number of clusters formed must generally be used for bi-directional weight movements to be beneficial.

2.3.5. Cluster relabeling

If the match function $S_J(I)$ is close to *1* (implying that the training point fell close to a very small weight rectangle), and yet the classes associated with cluster J and training vector I differ, the standard Fuzzy ARTMAP creates a new cluster. Since tiling even a small volume of the feature space with very small rectangles requires a large number of clusters, this is generally undesirable. Instead, it may be desirable to increment the MAP field entry $(J, class_I)$ if $S_J(I) > \rho_{relabel}$, where $\rho_{relabel}$ is a threshold constant close to *1*. This leads to two non-zero entries in the MAP field, associating one cluster with two classes. When a test vector is captured by such a cluster, it is possible to assign a the vector a fuzzy membership in each of the classes, or to output a single class label based on the relative magnitudes of the two MAP field entries (consistent with section 2.3.1). If cluster relabeling is utilized, it is important to also choose the choice function favoring larger weight rectangles (section 2.3.3) to prevent clusters associated with multiple classes from capturing points which can be unambiguously assigned to a single class.

2.4. Results

We have tested the above variants on several data sets, including artificial data, formant frequencies of steady state vowels (2 to 4 dimensional vectors), and cepstral coefficients extracted from continuous speech (10 to 20 dimensional vectors). On each data set, the best results were achieved using a combination of the proposed algorithm modifications. However, no combination of variants and parameter values yielded particularly good performance on all data sets. One cause of this is that the class distributions vary widely between data sets. Another is the varying dimensionality of the feature spaces – a fixed number of clusters will tile a twenty dimensional hypercube very differently than a two dimensional square.

Table 1 and Figure 3 illustrate the performance of the Fuzzy ARTMAP variants on the Peterson-Barney vowel data in two dimensions. Since several identical training data values are labeled with different classes, the number of clusters generated by the standard algorithm grows without bounds. The parameter set yielding the best classification generated 620 clusters after 20 iterations through the training set. The smallest number of clusters generated by the standard algorithm after one pass through the training data was 104, giving 57% correct classification of the test set. In contrast, the modified algorithm achieved up to 72% correct classification with 105 clusters and 74% correct with 141 clusters.

α	β	Number of iterations	Max. clusters per iteration	N_{prune}	ρ	$\rho_{relabel}$	Choice function	Bi-directional weight movement	Number of clusters	Percent correct
Modified Fuzzy ARTMAP:										
.0001	**0.1**	**20**	**20**	**3**	**0.5**	**-**	**Alternate**	**√**	**141**	**74**
.0001	0.01	5	-	-	0.5	0.99	Alternate		121	73
0.01	0.05	5	-	-	0.5	0.99	Alternate		96	72
0.01	0.05	5	-	-	0.5	0.99	Weber		105	72
0.01	0.05	20	-	3	0.5	-	Weber		260	72
.0001	0.05	40	20	5	0.5	-	Alternate	√	109	71
0.01	0.1	5	-	-	0.5	0.99	Alternate		87	69
Standard Fuzzy ARTMAP:										
0.01	**0.05**	**20**	**-**	**-**	**0.5**	**-**	**Weber**		**620**	**71**
0.01	0.05	5	-	-	0.5	-	Weber		450	69
0.01	1.0	5	-	-	0.5	-	Weber		243	68
0.01	1.0	2	-	-	0.5	-	Weber		181	66
0.01	0.25	5	-	-	0.5	-	Weber		247	65
0.01	1.0	1	-	-	0.5	-	Weber		104	57

Table 1 Performance of Fuzzy ARTMAP on the Peterson-Vowel data (formant frequencies f_1 and f_2). The parameter values shown here yielded the highest performance. Results of numerous simulations yielding equivalent or worse performance are not included. The **bold** parameters were used to generate the plots in Figure 3.

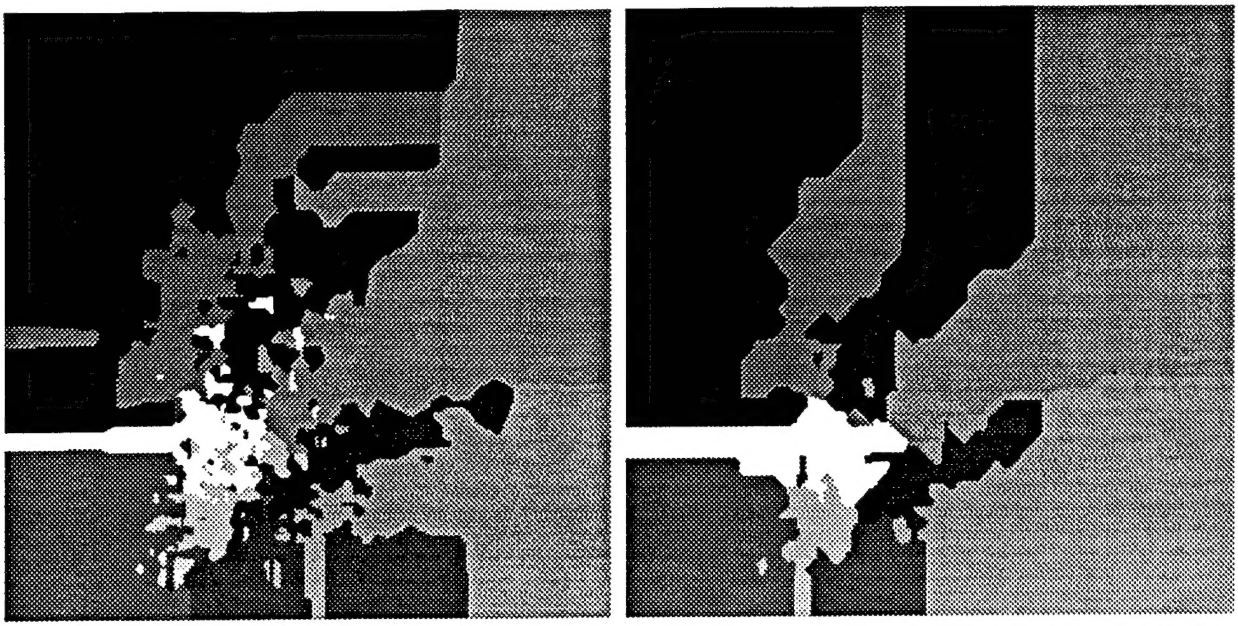

Figure 3 Map of feature space generated by the standard Fuzzy ARTMAP algorithm, left, and by the modified algorithm using bi-directional weight movement, limited cluster creation, pruning, and the alternate choice function, right. The numerous small clusters in the left plot indicate overtraining by the standard algorithm. Both variants were trained on 1020 points randomly chosen from the Peterson-Barney vowel data plotted in Figure 1.

Note that this is a very difficult classification task – for comparison, the best classification accuracy we were able to achieve with optimized-learning-rate LVQ1[7] initialized by k-means clustering was 70.2% correct. While the classification performance of LVQ is similar to that of the modified Fuzzy ARTMAP, the latter provides several advantages. Notably, due to the use of the L_1 norm, Fuzzy ARTMAP does not require a multiplication operation per input vector component. Furthermore, it need not be initialized using an unsupervised clustering algorithm, facilitating on-line learning. Finally, because the Fuzzy ART weights directly define not only the position, but also the extent of each cluster, similar classification accuracy can sometimes be achieved with fewer clusters.

We presented this work to at the World Congress on Neural Networks in September, 1996.

3. Optical transpose interconnection system

Optical interconnections have advantages over electrical interconnections in terms of speed, power, and density for high-speed links over distances on the order of centimeters[8]. Here, we are concerned with a particular interconnection pattern, the *k*-shuffle, which is commonly used in switching applications and to interconnect processors in parallel computing systems[9]. Section 3.1 shows how to map the backpropagation neural network onto this interconnection pattern. The optical transpose interconnection system[10] (OTIS) architecture provides the *k*-shuffle using only a pair of lenslet arrays. We designed, fabricated, and tested birefringent computer generated holograms (BCGH) for implementing this system with reflective modulators. Furthermore, we developed a novel optical system for implementing the OTIS architecture utilizing surface normal transmitters. We carried out a detailed analysis and optimization of the latter system based on Gaussian beam propagation theory.

The k-shuffle is a one-to-one interconnection between L input nodes and L output nodes, where L is the product of two integers, M and N. The input nodes are divided into M groups of N members, and the output nodes are divided into N groups of M members. The k-shuffle connects input node (i, j) to output node (j, i), where the first number indexes the group and the second number indexes the element within a group. This operation may also be viewed as the transposition of an MxN matrix.

In the OTIS system, the nodes of each group are arranged in a square pattern to limit the length of electrical interconnections within the group. Similarly, the groups are arranged in a square pattern. Thus, we define $M = \#_{m1}\text{x}\#_{m1} = \#_{g2}\text{x}\#_{g2}$ and $N = \#_{m2}\text{x}\#_{m2} = \#_{g1}\text{x}\#_{g1}$, where $\#_m$ is the number of members along each edge of a group and $\#_g$ is the number of groups along an edge of the processing plane. OTIS uses one lenslet per group, as shown in cross-section in Figure 4.

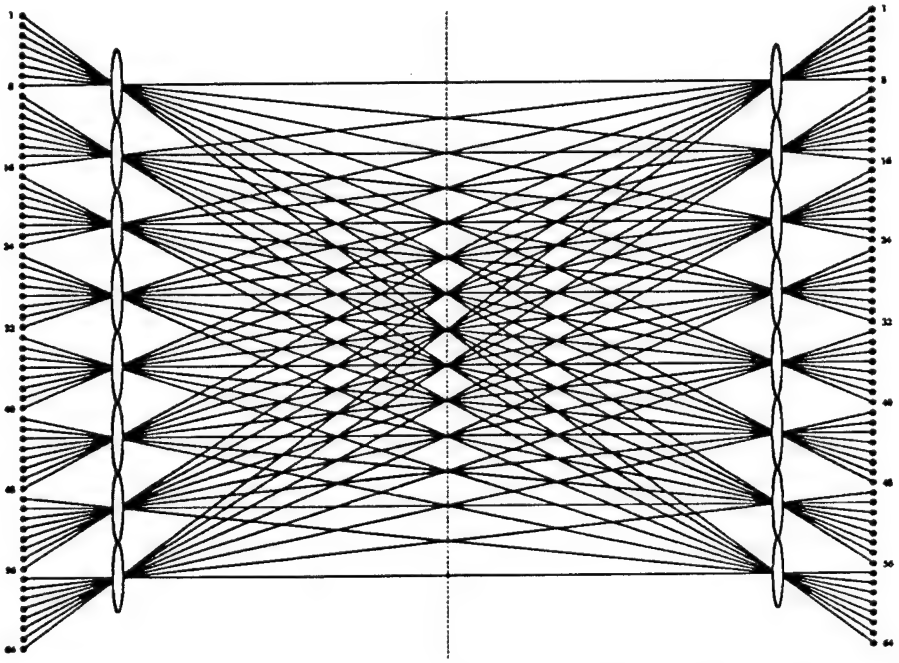

Figure 4. Optical transpose interconnection system for individually directed transmitters. This example is drawn for 8x8 groups of 8x8 transmitters (receivers) per group, for a total of 4096 channels.

3.1. Application to backpropagation

Currently, the most widely used neural network learning algorithm is probably backpropagation[11]. Here, we demonstrate how the OTIS system may be used to optically implement the interconnections between backpropagation processing units.

Figure 5 shows a neural architecture[12] for implementing the backpropagation algorithm. Each layer consists of a number of *units*. In order to conform to the requirement that a neural network processing element (PE) have only a single output, each unit is subdivided into a *sun* and multiple *planet* PEs. One backpropagation weight, w_{lij}, is stored at each planet.

During the forward pass, the output, z_{li}, of unit i of layer l is computed as the scaled sum of the outputs of the previous layer multiplied by the unit's corresponding weights:

$$z_{li} = \sigma\left(\sum_j w_{lij} z_{(l-1)j}\right) \equiv \sigma(I_{li})$$

During the backward pass, a set of error signals is calculated. Each unit's contribution, δ_{li}, to the global error is computed as the scaled sum of the errors of the next layer's error signals times the weights corresponding to that unit in the next layer:

$$\delta_{li} = \sigma'(I_{li}) \Sigma_k w_{(l+1)ki} \delta_{(l+1)k}$$

Each weight is modified in proportion to its input and error signals in order to reduce the global error.

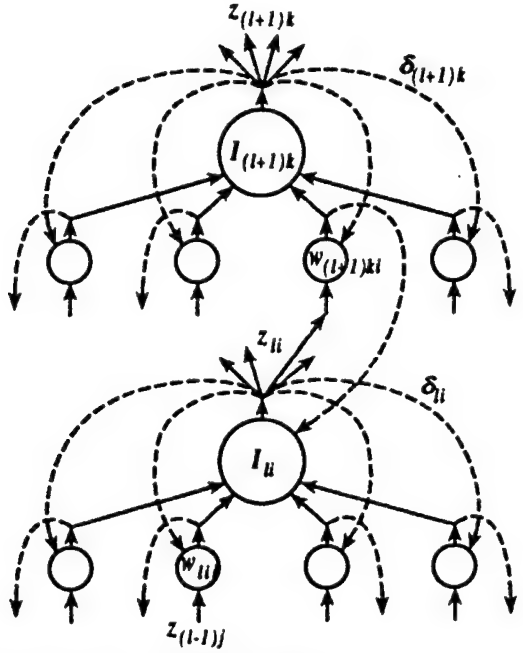

Figure 5. Two *units* of a backpropagation neural network.

This architecture may be mapped onto the OTIS optical system by devoting one OTIS group to each backpropagation unit and assigning a single transmitter/receiver pair to each planet. The interconnections within a unit are implemented electrically. The forward-pass communication between layers is implemented by electrically fanning out each sun's output to the associated planets and then transmitting the signals to the next layer utilizing the one-to-one optical interconnections. During the backward pass, the error signals, calculated at the planet PEs, are optically transmitted to the previous layer and electrically summed there. Thus, the long interconnections are implemented optically, whereas the short interconnections are implemented electrically.

A multi-layer backpropagation system could be implemented as multiple OTIS systems interfaced by beam splitters or as a single OTIS system with the electronics multiplexed to compute the signals for the multiple layers sequentially.

3.2. System demonstration

For high efficiency and power uniformity, the optical system shown in Figure 4 requires individually directed transmitters. To achieve directed signal beams in a system utilizing reflective modulators (Figure 6), an additional array of illumination lenslets is required (Figure 7). However, the interconnect lenslets and illumination lenslets may be implemented as a single optical element if birefringent computer-generated hologram (BCGH) technology[13] is utilized. We have designed, built, and characterized such holograms. One of the BCGH elements is shown in Figure 9. The modulator input spots generated by this element are shown in Figure 11, and the spots produced by the same element at the intermediate image plane of the

OTIS interconnection system are shown in Figure 12. Beam scans through these planes are shown in Figure 13. The spot size measured using the beam scan instrument tends to be less accurate (larger) than that measured by imaging the spots onto a CCD camera because of the difficulty in aligning the BeamScan's slit with a row of spots.

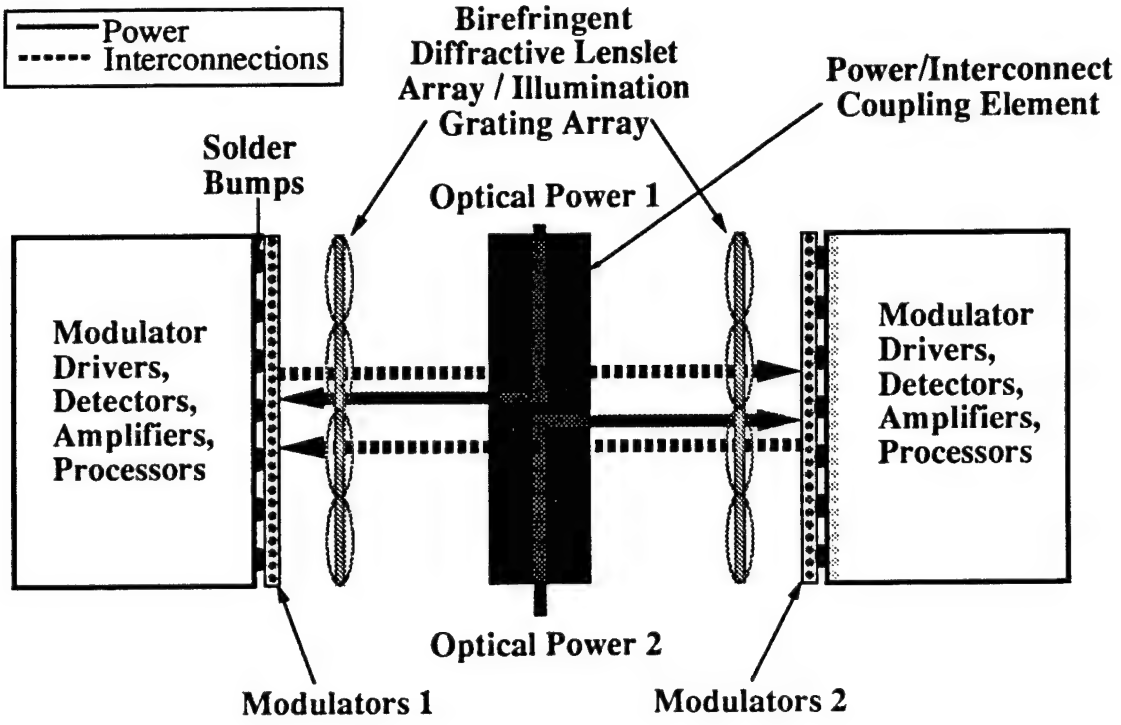

Figure 6. Schematic diagram of an optoelectronic system utilizing reflective modulators.

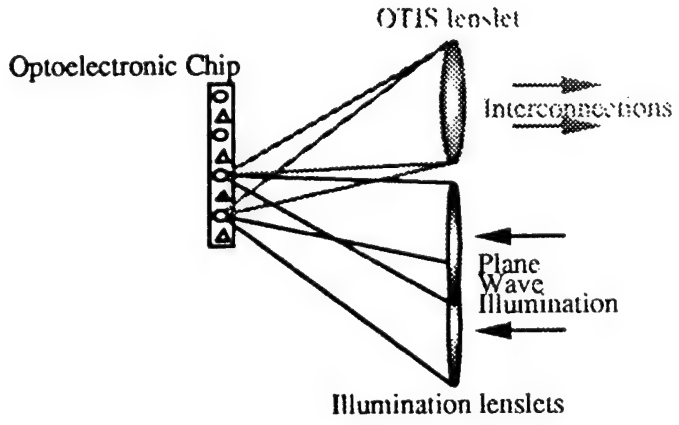

Figure 7. Illumination lenslets and interconnect lenslets may be superimposed in the same plane by utilizing birefringent computer-generated holograms (BCGH).

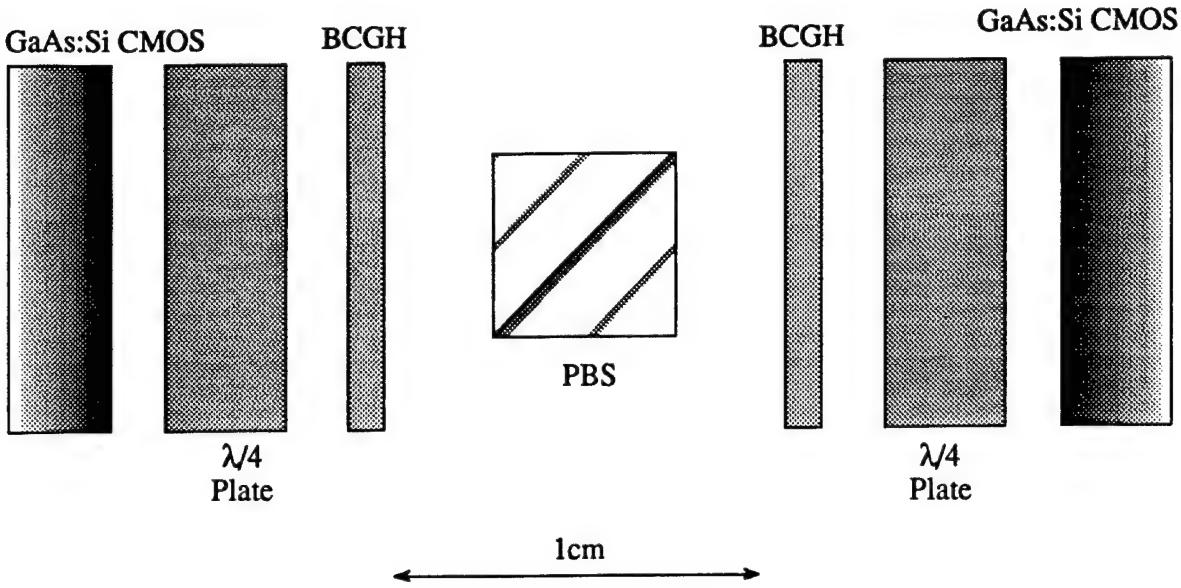

Figure 8. OTIS demonstration system schematic. The experimental system bi-directionally interconnects two planes, each consisting of an 8x4 modulator array and an 8x4 detector array. The system length is 2.48 centimeters.

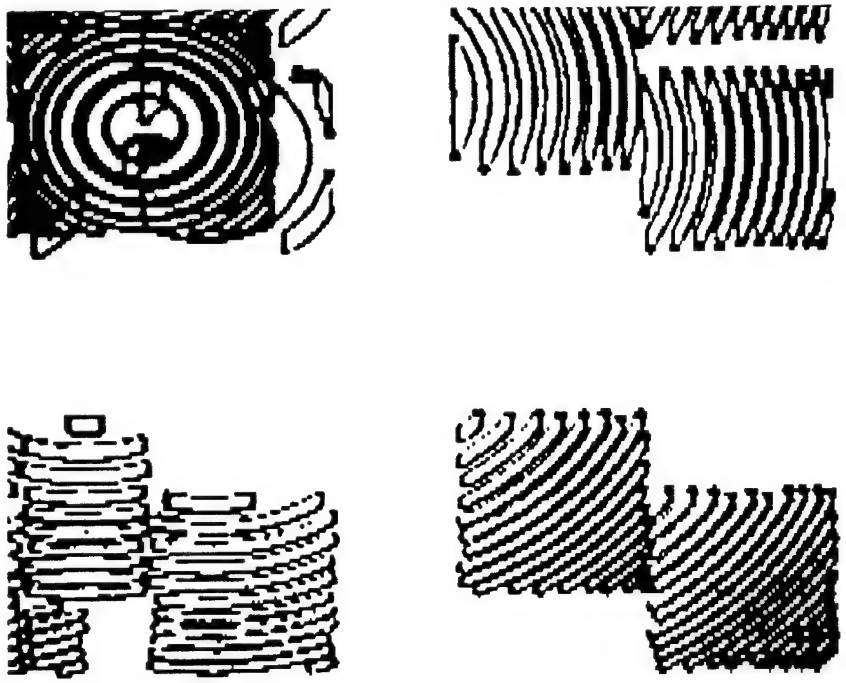

Figure 9. Birefringent computer generated hologram implementing both the illumination lenslets and the OTIS interconnection lenslets. The minimum feature size of the hologram is 5μm.

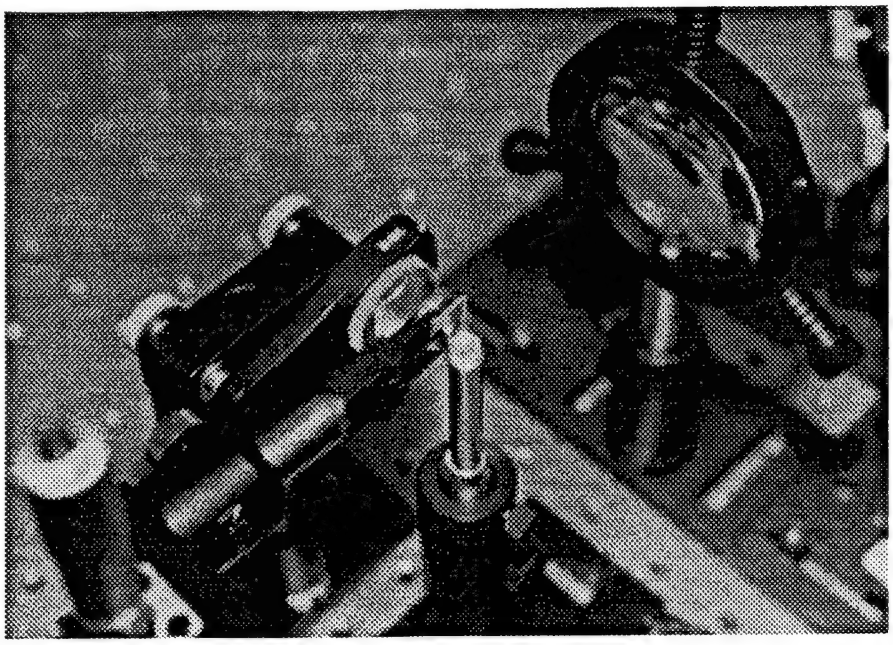

Figure 10. Optical bench set-up used to characterize the OTIS BCGH.

Figure 11. Modulator input generated by BCGH illumination lenslets. The full width at half-maximum (FWHM) spot diameters in the modulator plane range from 25μm to 37μm.

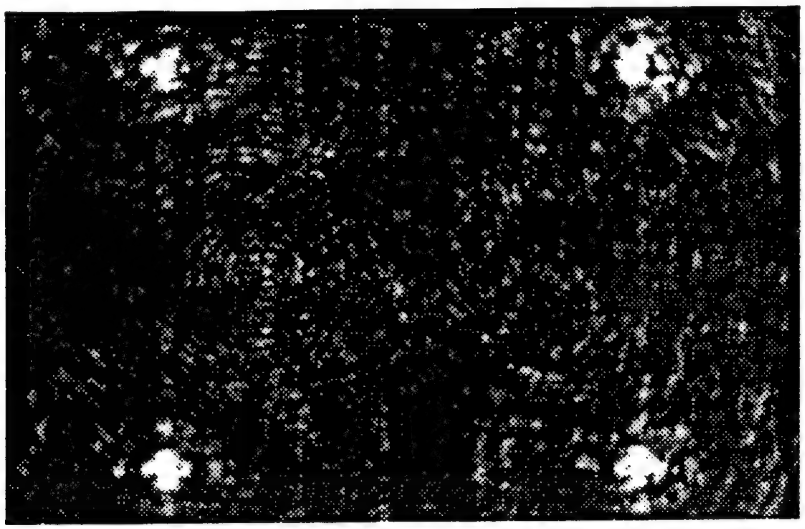

Figure 12. Spots in the intermediate image plane (the focal plane of the interconnect elements) of the OTIS system. The light recorded here has passed through the BCGH elements twice, experiencing a different lens function on each of the two passes.

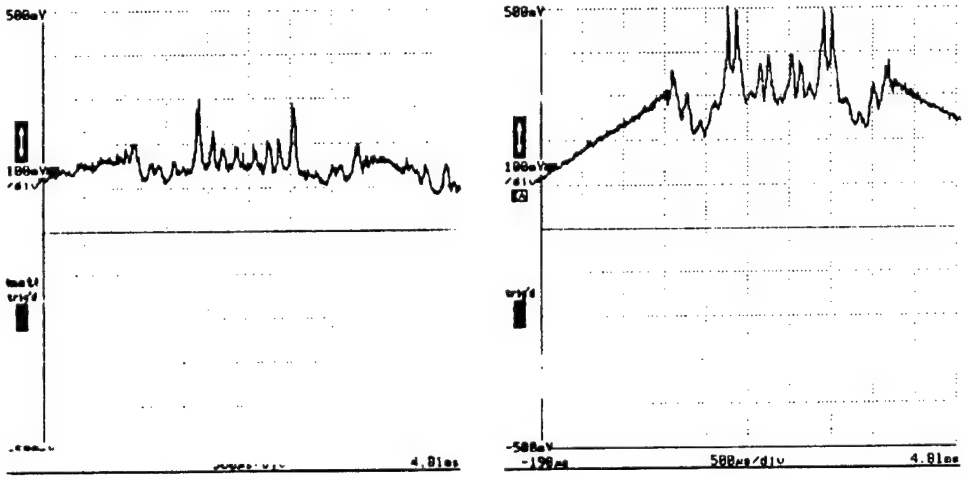

Figure 13. Beam scans of spots in the modulator plane (left) and intermediate image plane (right) of the OTIS system.

3.3. Optical transpose interconnection system for vertical emitters (OTIS-VE)

We developed a novel optical system design for implementing the OTIS architecture utilizing surface normal transmitters. This eliminates the need for BCGH (described in the previous section) to implement a reflective modulator-based system, and facilitates implementations utilizing vertical cavity lasers (VCSELs). The drawback of the newer design is that the system length is in some cases longer than that of the original design. Which system is most suitable for a given application will depend on the type of transmitters used, the desired spacing between transmitters and between groups, available packaging technology, and the relative penalties for crosstalk, optical losses, and system size.

In the OTIS system, each lens axis is generally not centered over a group. However, if surface-normal transmitters are to be used, each lens aperture should be centered over the corresponding group. Thus, the lens functions are asymmetrically cropped (Figure 14). The parallel beams from a group of transmitters

cross at the rear focal point of the lens corresponding to that group. Each beam then propagates through the focal point of a lens in the second lenslet plane. In addition to focusing the beam, the second lenslet deflect the beam to again be parallel to the system axis.

The optical system is bi-directional. Furthermore, if $\#_{g1}=\#_{g2}$ and $\#_{m1}=\#_{m2}$, a folded system may be constructed using a single lenslet array and a mirror to interconnect nodes within a single processing plane.

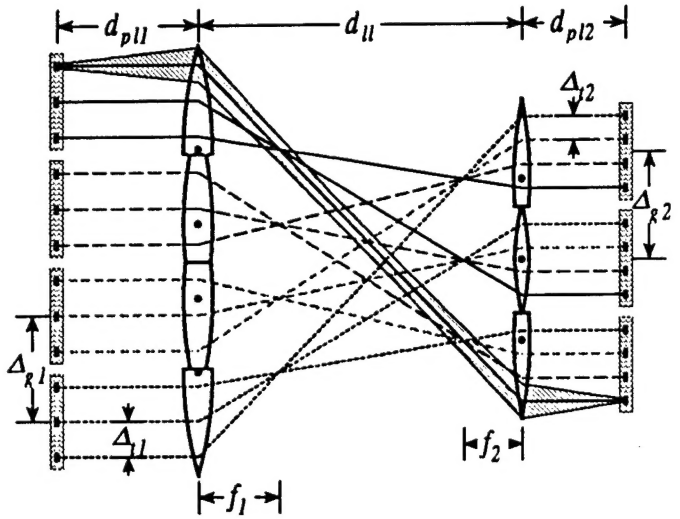

Figure 14. Optical transpose interconnection system for vertical emitters (OTIS-VE). Two planes of lenslets rearrange the data such that each group in the left plane is connected to all groups in the right plane and visa versa. The system is bi-directional.

3.4. Gaussian beam analysis

We carried out a Gaussian beam analysis[14] in order to determine the feasibility of using current refractive or diffractive (CGH) optics technology to build the system described above.

The basic equations governing the propagation of Gaussian beams are propagation of a beam in free space:

$$\omega^2(z) = \omega_0^2\left(1+\frac{z^2}{z_0^2}\right), \quad R(z) = z\left(1+\frac{z_0^2}{z^2}\right), \quad z_0 \equiv \frac{\pi\omega_0^2}{\lambda}$$

and propagation through a thin lens:

$$\frac{1}{R_2} = \frac{1}{R_1} - \frac{1}{f},$$

as illustrated in Figure 15 and Figure 16.

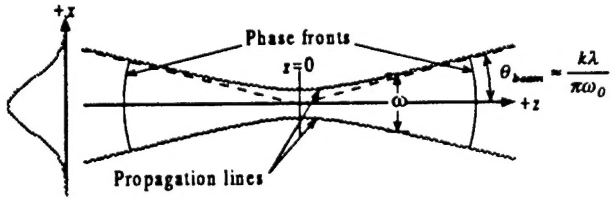

Figure 15. Gaussian beam propagation in free space.

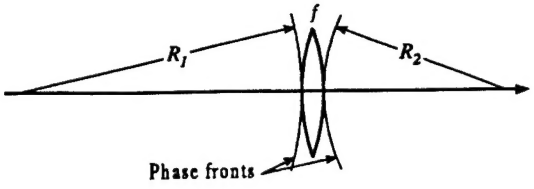

Figure 16. Effect of a thin lens.

The geometry of the interconnection (similar triangles in Figure 14) leads to the relationship

$$d_{ll} = \left(\Delta_{g2}/\Delta_{t1} + 1\right) f_1 = \left(\Delta_{g1}/\Delta_{t2} + 1\right) f_2 .$$

Furthermore, the constraint is imposed that none of the beams may be clipped (by a lens aperture) at a radius less than k times the radius of the *e-2* irradiance contour. A value of k=2.12 assures that the beams remain Gaussian[15] and corresponds to clipping less than 0.1% of the optical power. The maximum acceptable radius of the *e-2* irradiance contour at the detector is specified as ω_{det}. Finally, the technology parameter F_{min} indicates the smallest permissible *f*/number of the lenslets.

These equations lead to a constrained search of the five-dimensional parameter space Δ_{g1}, Δ_{g2}, f_1, d_{pl1}, d_{pl2}, the objective being to find the shortest length system that satisfies the above requirements. Since the system is bi-directional, the clipping constraints are checked at both lenslet planes for beams propagating in both directions. Typical plots of minimum system length vs. input beam radius are shown in Figure 17 for a symmetrical system consisting of 64x64 transmitters in each plane. This analysis shows that the system is buildable based on reasonable values of the technology parameters.

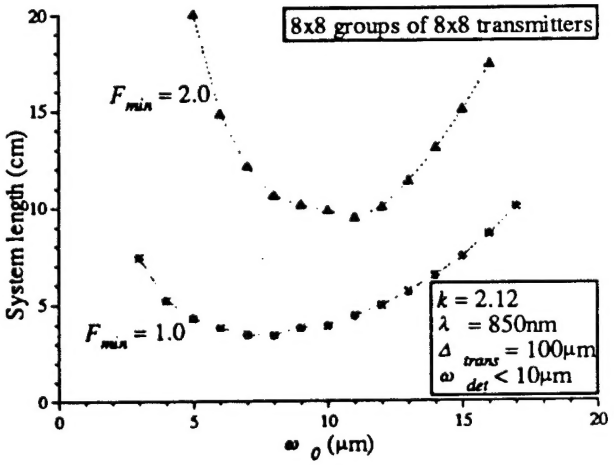

Figure 17. Length of a system with 64x64 transmitters in each plane for various values of the input beam waist, ω_0, and the minimum achievable f/number, F_{min}.

The plots show that it is possible to simultaneously achieve de-magnification of beams propagating in both directions. While the optical system is symmetrical, the beams are not – a beam waist occurs at the input plane but not necessarily at the output plane.

For small values of ω_0 the *f*/number constraint causes the system to be large or not feasible. For example, for a wavelength of λ=0.85μm, far from the beam waist, an *f*/2 lens can just capture a beam with ω_0=2.3 without cropping it inside the k=2.12 contour. It can not capture several such beams that are adjacent to one another without excessively clipping some of them. For large values of ω_0, the input beams are nearly collimated. In this case, their intermediate beam waists occur near the rear focal points of the first lenslets. Consequently, the beams expand significantly before reaching the second lenslet plane. Thus, the second

lenslet apertures must be large, again resulting in a large system. The combination of these two effects leads to an optimal input beam size that results in the smallest optical system.

In summary, the theoretical optical efficiency of the novel OTIS system for surface-normal transmitters is nearly 100%. The Gaussian beam analysis shows that fabrication of the system is feasible in today's technology.

We have submitted a paper describing this work to the OSA topical meeting on Optics in Computing, to be held in March, 1997.

4. Major accomplishments

- Refinement of a the Fuzzy ARTMAP neural network algorithm. The algorithm is very well suited to hardware implementation, but previously did not perform well on most real problems.
- Design, fabrication, and characterization of birefringent computer generated holograms to implement the OTIS optical system for reflective modulators. This optical system provides a *k*-shuffle interconnection pattern that may be used to implement a backpropagation neural network.
- Development of a simple, compact, efficient optical system (OTIS-VE) compatible with surface-normal transmitters (such as VCSELs) implementing the *k*-shuffle interconnection pattern.

5. Future work

During this project, we focused our efforts on the development of hybrid optoelectronic neural modules, as well as on defining guidelines as where to use optoelectronics in neural computing hardware. Several additional projects are currently under way to demonstrate and characterize system prototypes utilizing this technology.

6. References

[1] M. Blume and S. C. Esener, "An efficient mapping of Fuzzy ART onto a neural architecture", accepted for publication in *Neural Networks*, August, 1996.

[2] Peterson, G. E. & Barney, H. L., "Control methods used in a study of the vowels", *Journal of the Acoustical Society of America*, 24(2), p. 175-184, 1952.

[3] Carpenter, G. A., Grossberg, S., & Rosen, D. B., "Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system", *Neural Networks*, 4(6), p. 759-771, 1991.

[4] Carpenter; G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B., "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps", *IEEE Transactions on Neural Networks*, 3(5), p. 698-713, 1992.

[5] Kohonen, T., Barna, G. & Chrisley, R., "Statistical pattern recognition with neural networks: benchmarking studies", in *IEEE International Conference on Neural Networks*, **1**, p. 61-68, IEEE, Piscataway, NJ, 1988.

[6] Bezdek, J., *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, 1981.

[7] Kohonen, T., Kangas, J., Laaksonen, J., & Torkkola, K., "LVQPAK: a software package for the correct application of Learning Vector Quantization algorithms", in *IJCNN International Joint Conference on Neural Networks*, **1**, p. 725-730, IEEE, New York, 1992.

[8] A. V. Krishnamoorthy, P. J. Marchand, F. E. Kiamilev, and S. C. Esener, "Grain-size considerations for optoelectronic multistage interconnection networks", *Applied Optics*, **31**:26, p. 5480-5507, 1992.
[9] F. T. Leighton, *Introduction to parallel algorithms and architectures*, Morgan-Kaufmann, San Mateo, CA, 1992.
[10] G. C. Marsden, P. J. Marchand, P. Harvey, and S. C. Esener, "Optical transpose interconnection system architectures", *Optics Letters*, **18**:13, p. 1083-1085, 1993.
[11] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation", in D. E. Rumelhart and J. L. McClelland [Eds.], *Parallel Distributed Processing*, **I**, p. 318-362, MIT Press, Cambridge, MA, 1986.
[12] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley, Reading, MA, 1989.
[13] F. Xu, J. E. Ford, and Y. Fainman, "Polarization-selective computer-generated holograms: design fabrication, and applications", *Applied Optics*, **34**, p. 256-266, 1995.
[14] D. O'Shea, *Elements of Modern Optical Design*, John Wiley & Sons, New York, 1985.
[15] F. B. McCormick, et al., "Optical interconnections using microlens arrays", *Optical and Quantum Electronics*, **24**:4, p. S465-S477, 1992.